

FIG 1

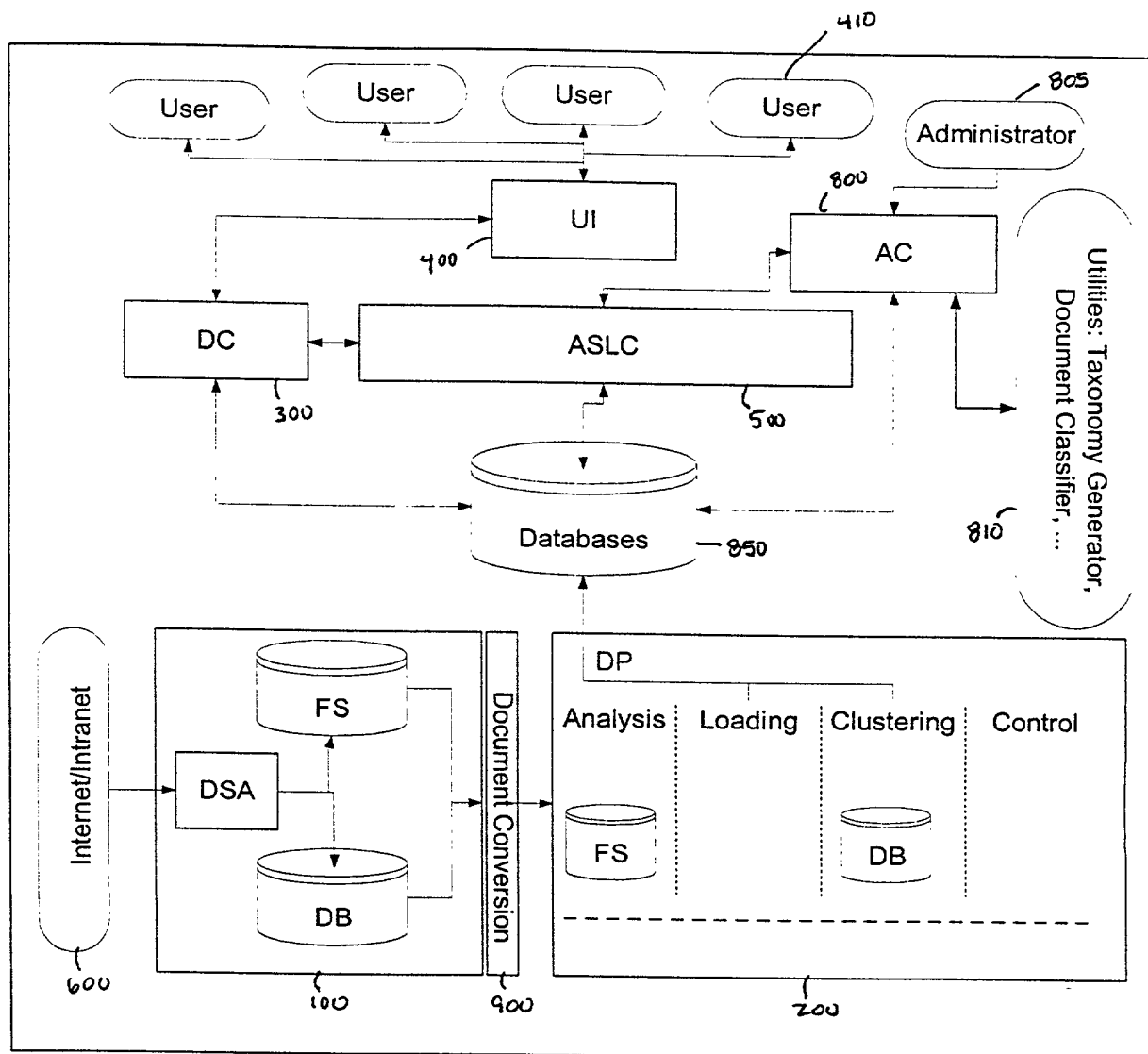


FIG 2

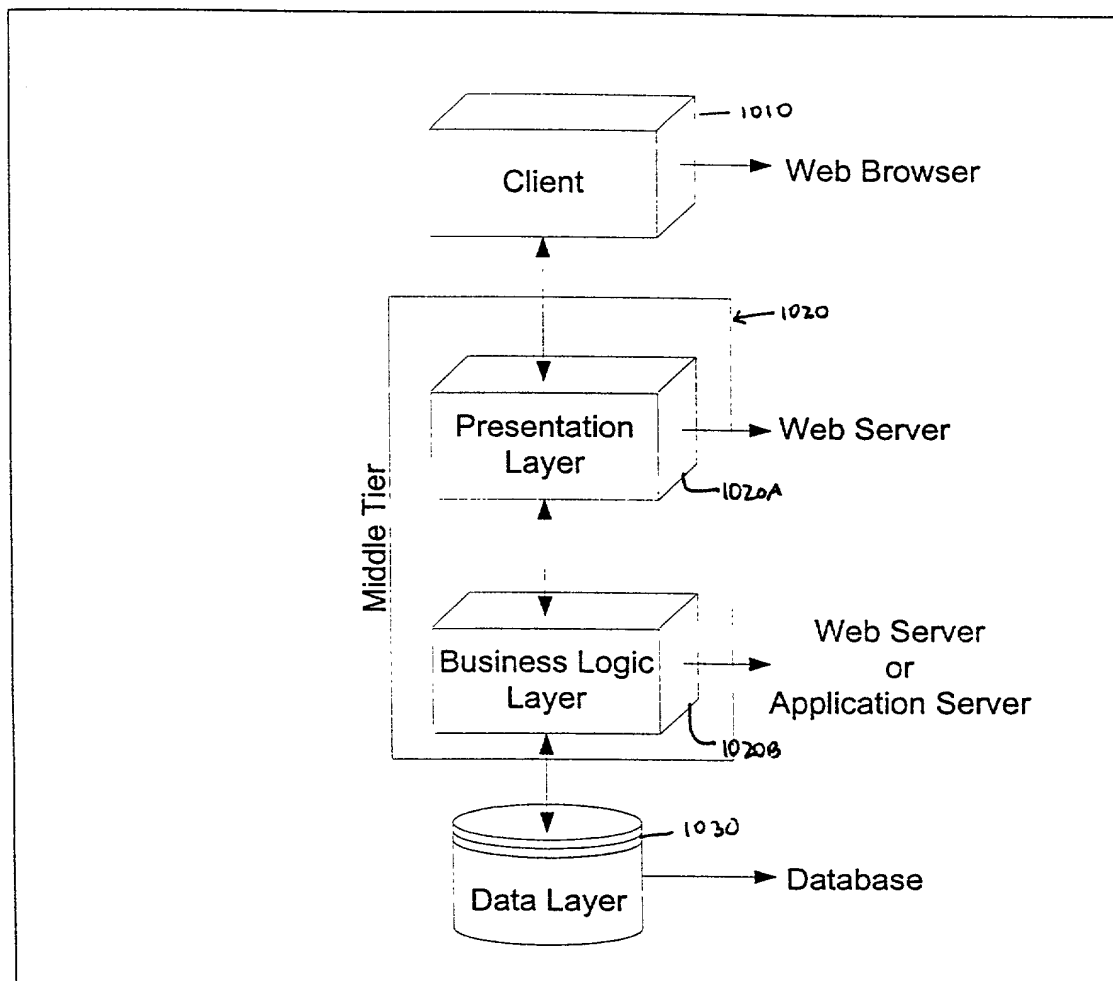


FIG 3

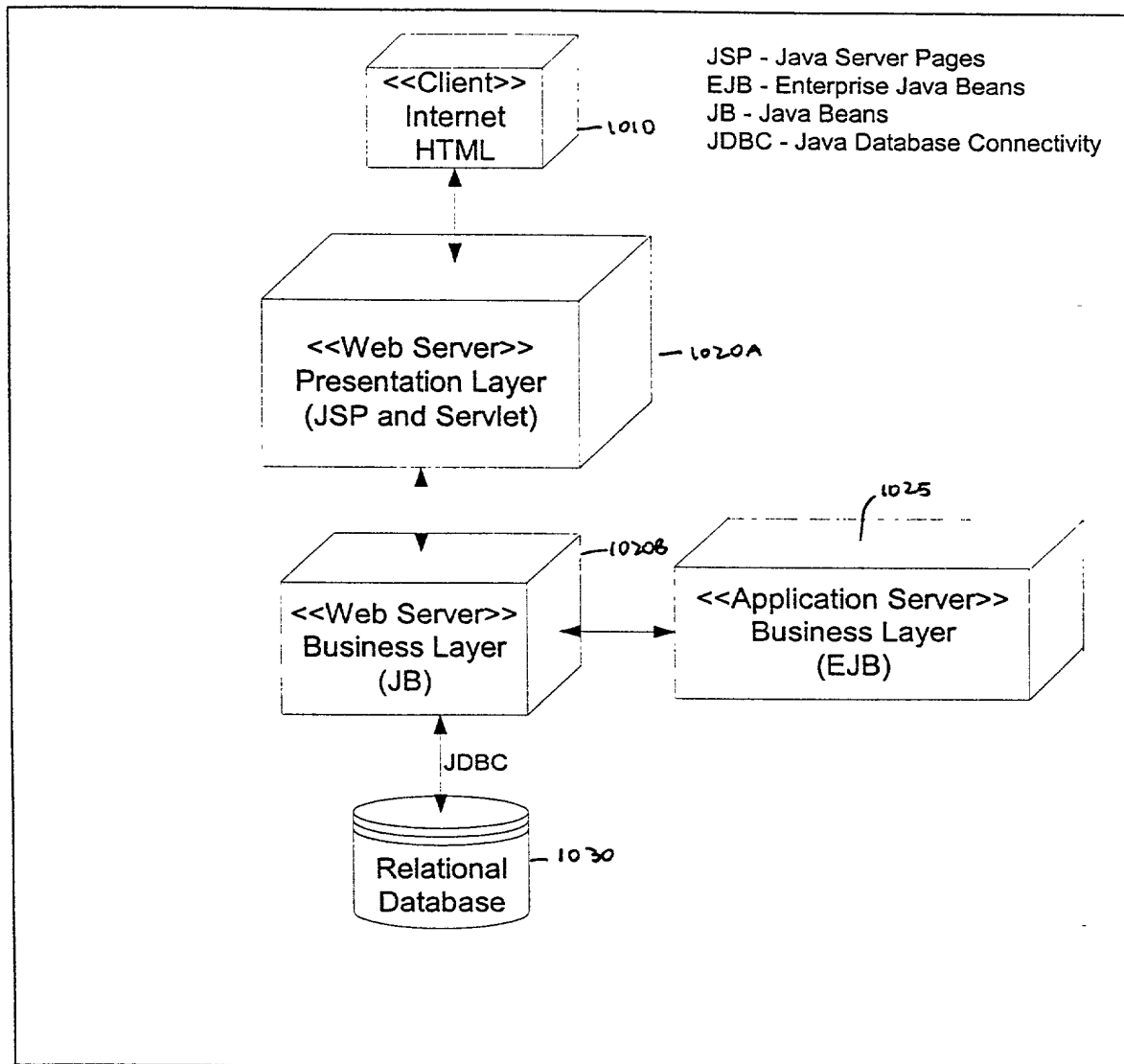


FIG 4

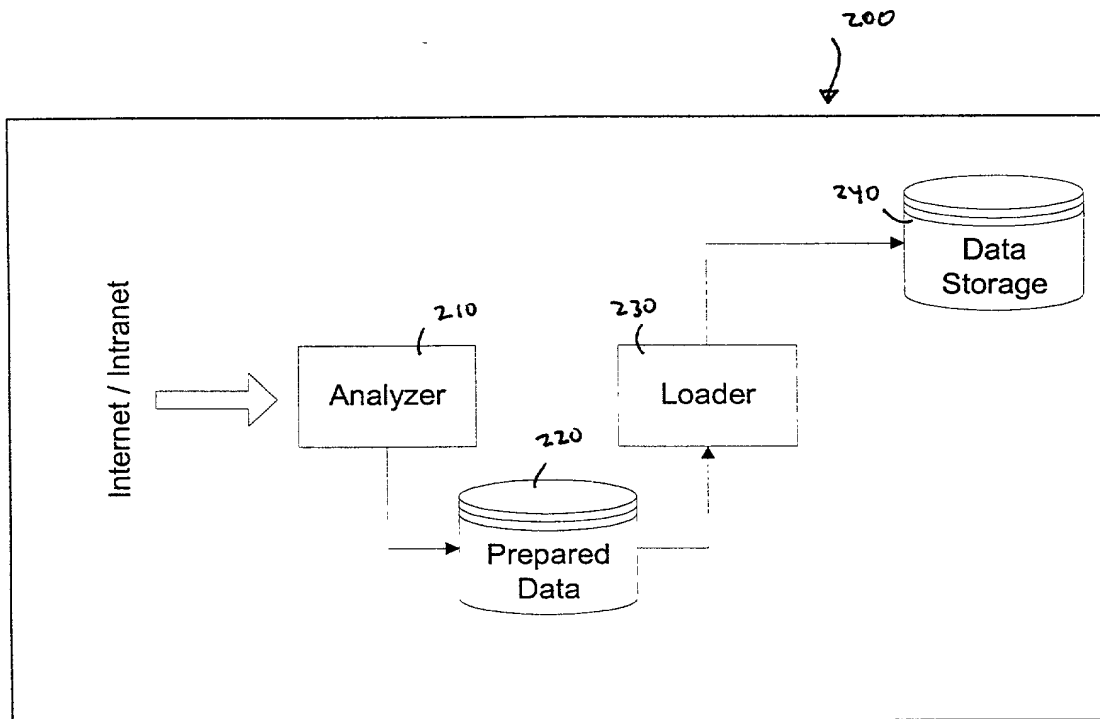


FIG 5

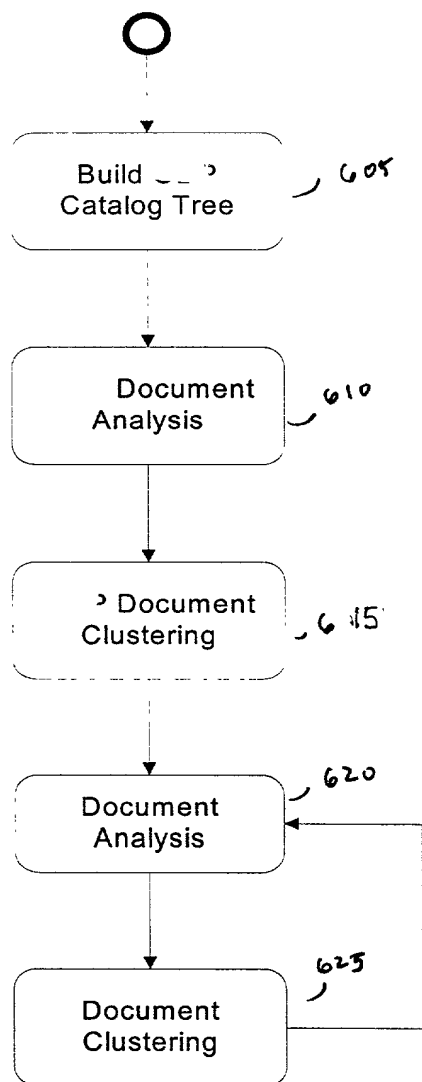


FIG 6

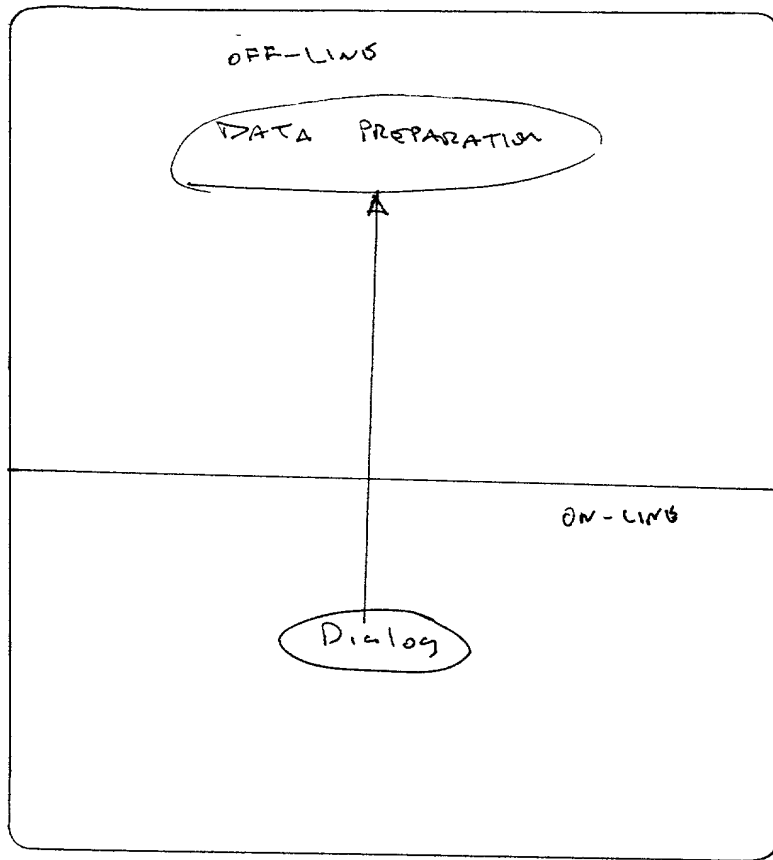


Figure 7

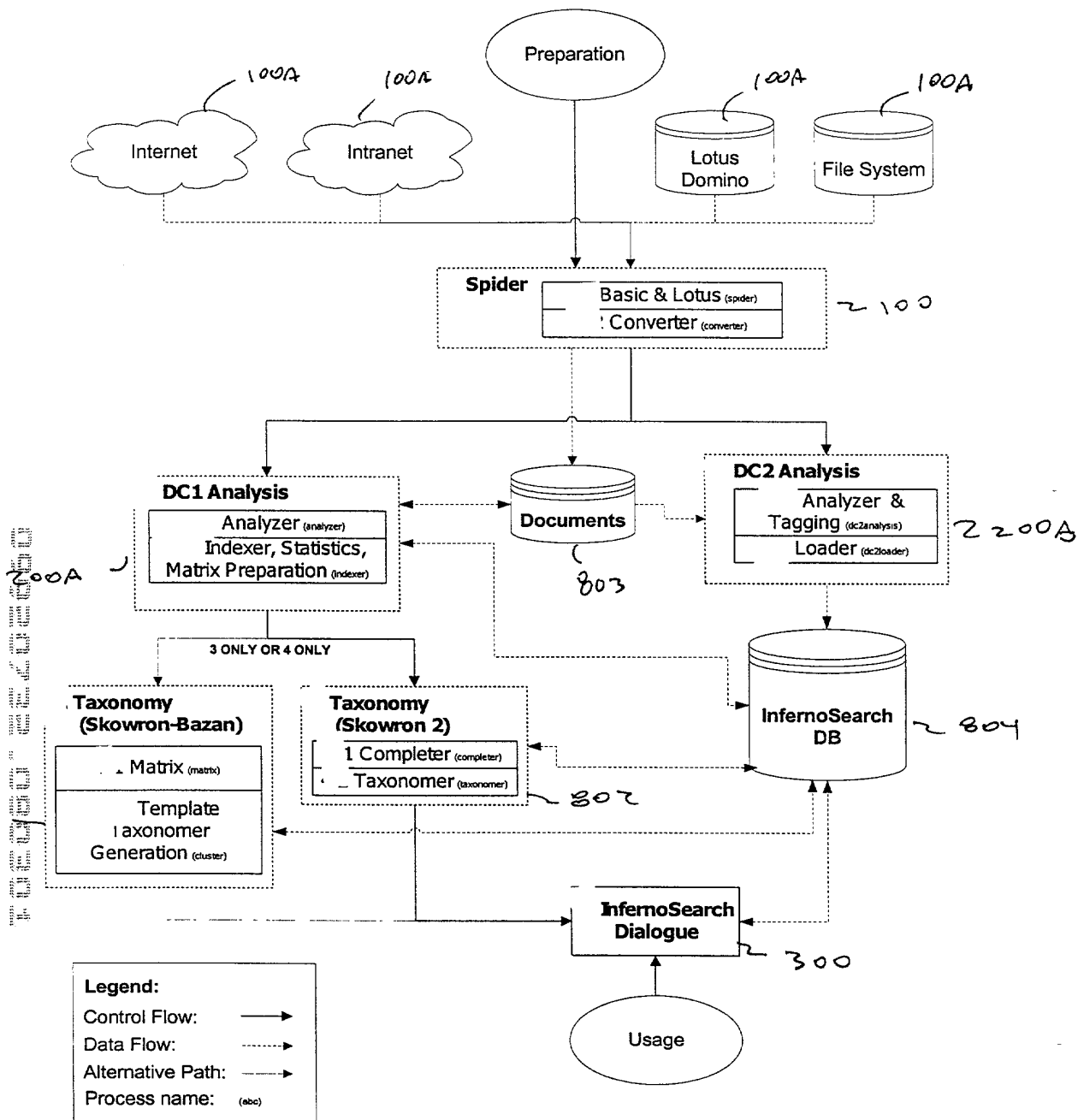
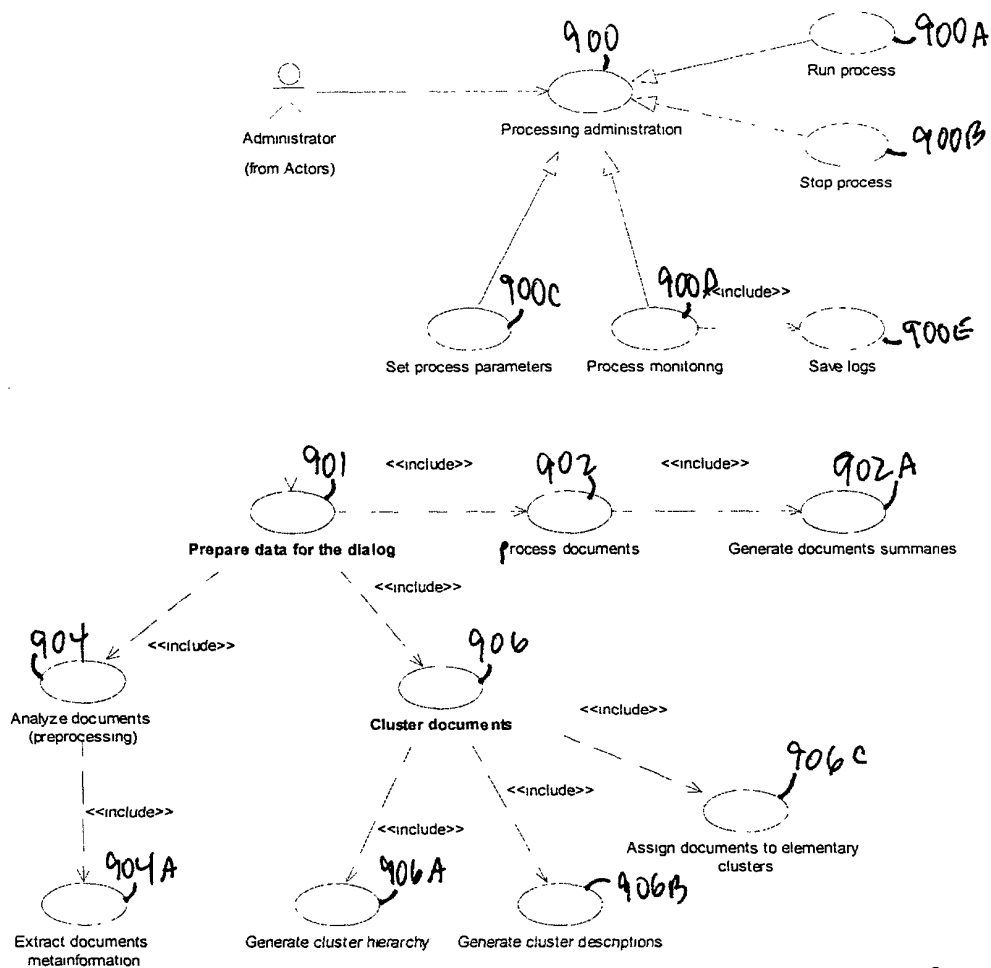


FIGURE 2



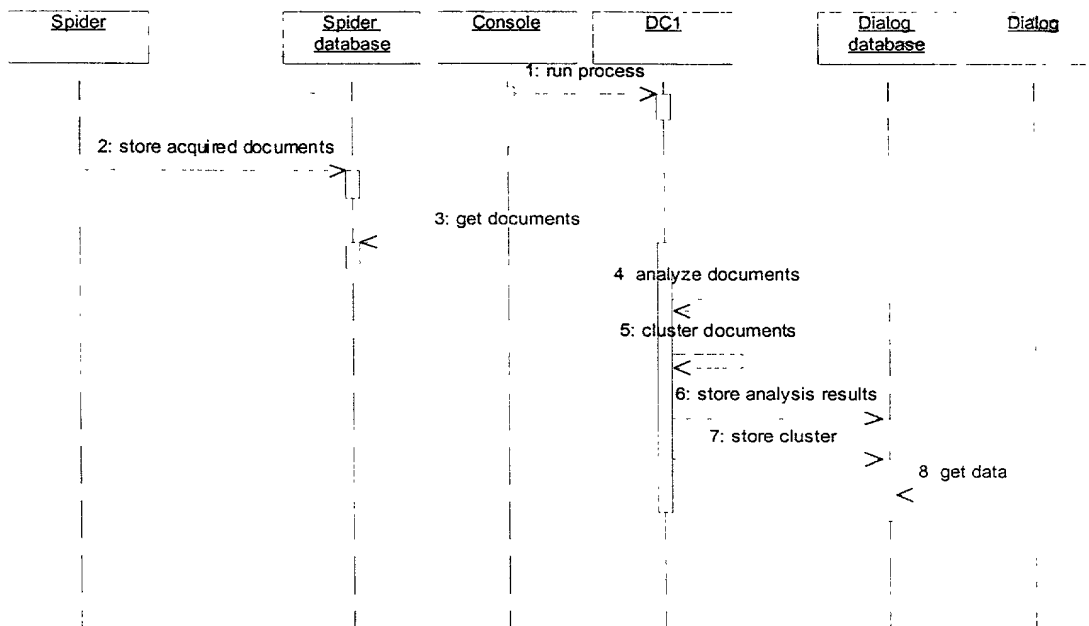


Figure 10

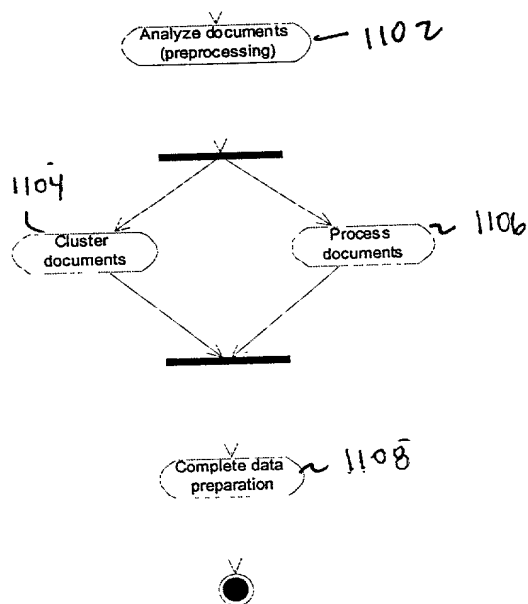


Fig. 11

Figure 12

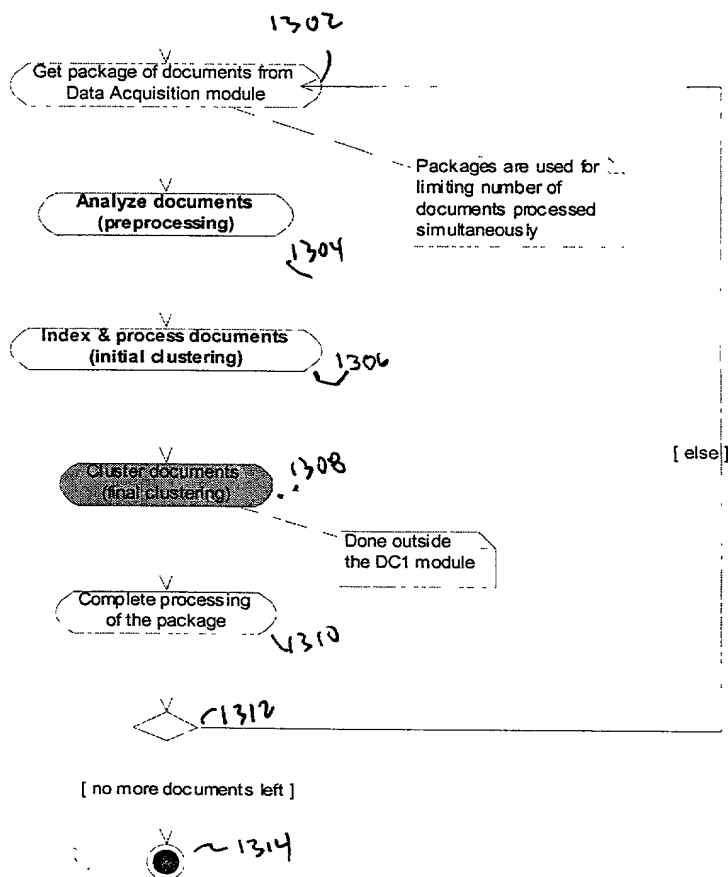
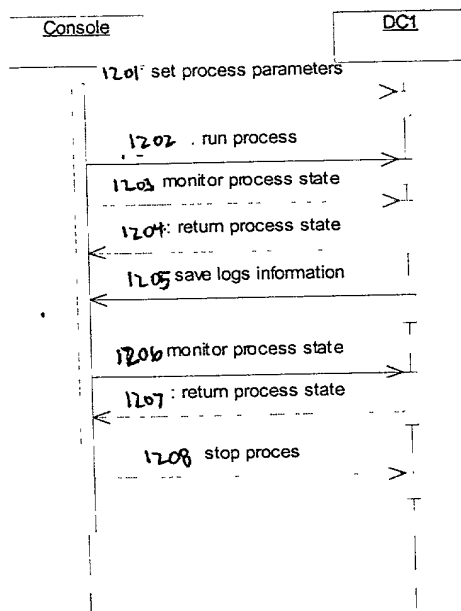


Figure 13

Figure 14

Terms "cluster" and "set of documents containing the same phrase" are equivalent on this diagram

Create local reverted index and dictionary of words/phrases - 1502

Generate documents' summaries and prepare statistical data for final clustering - 1504

Update global data with local data - 1506

Generate clusters hierarchy - 1508

Generate clusters descriptions - 1510

Find elementary clusters with too many documents - 1512

Find documents not covered by any elementary cluster - 1514

Complete initial clustering - 1516

Figure 15

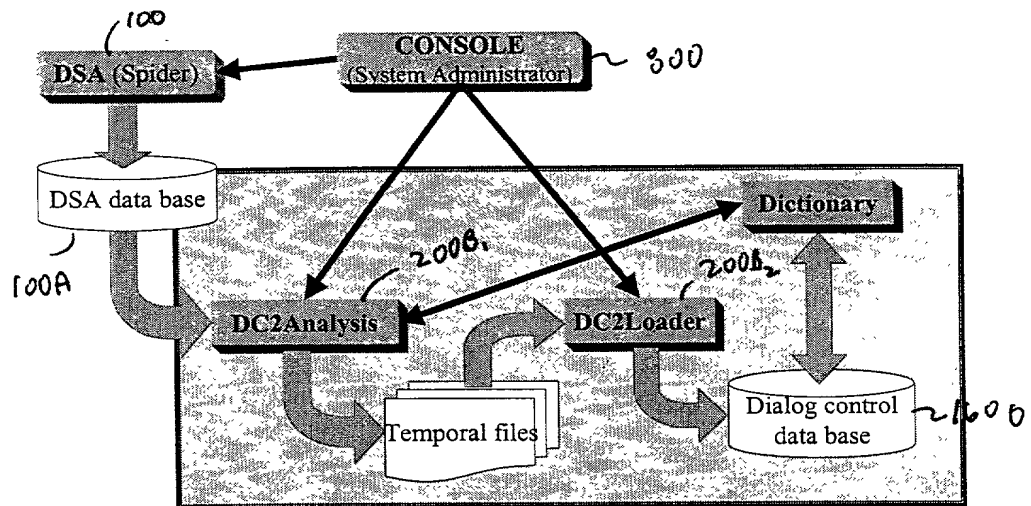


Figure 16

Legends

	Data flow
	Control flow

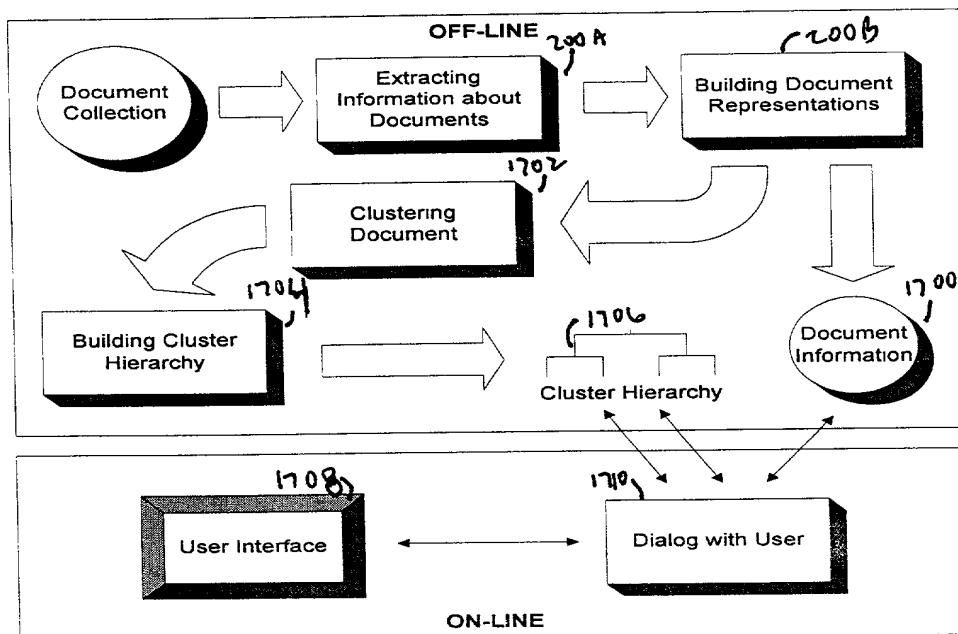


Figure 17

Copyright 2000 by International Business Machines Corporation. All rights reserved. IBM, the IBM logo, and the e-business logo are trademarks of International Business Machines Corporation. Other names may be trademarks of their respective owners.

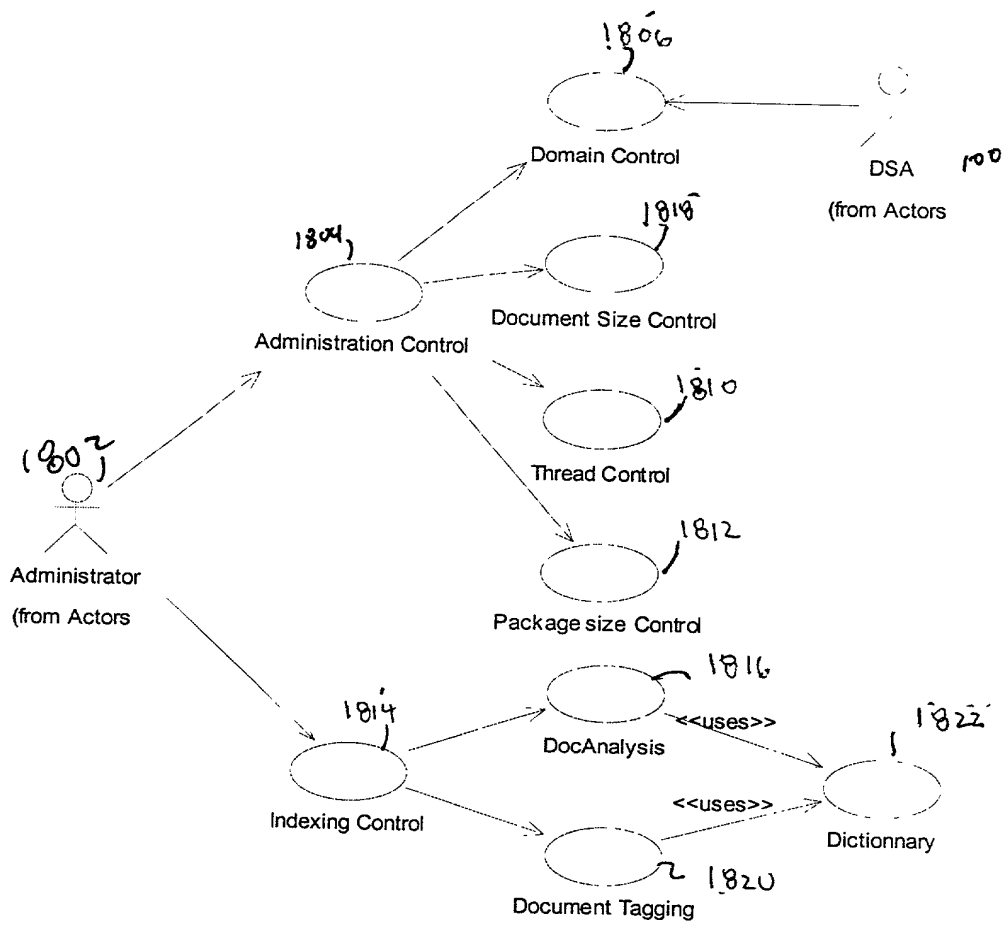


Figure 18

●

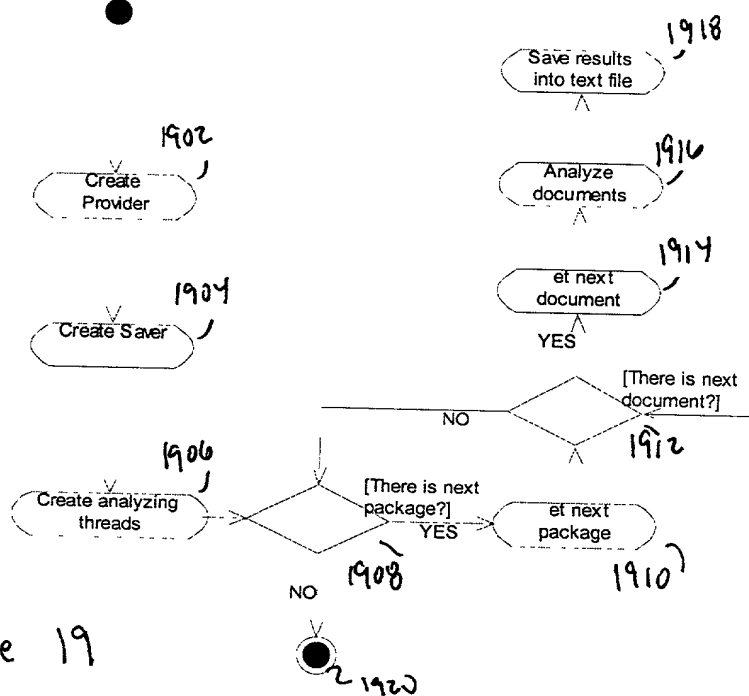


Figure 19

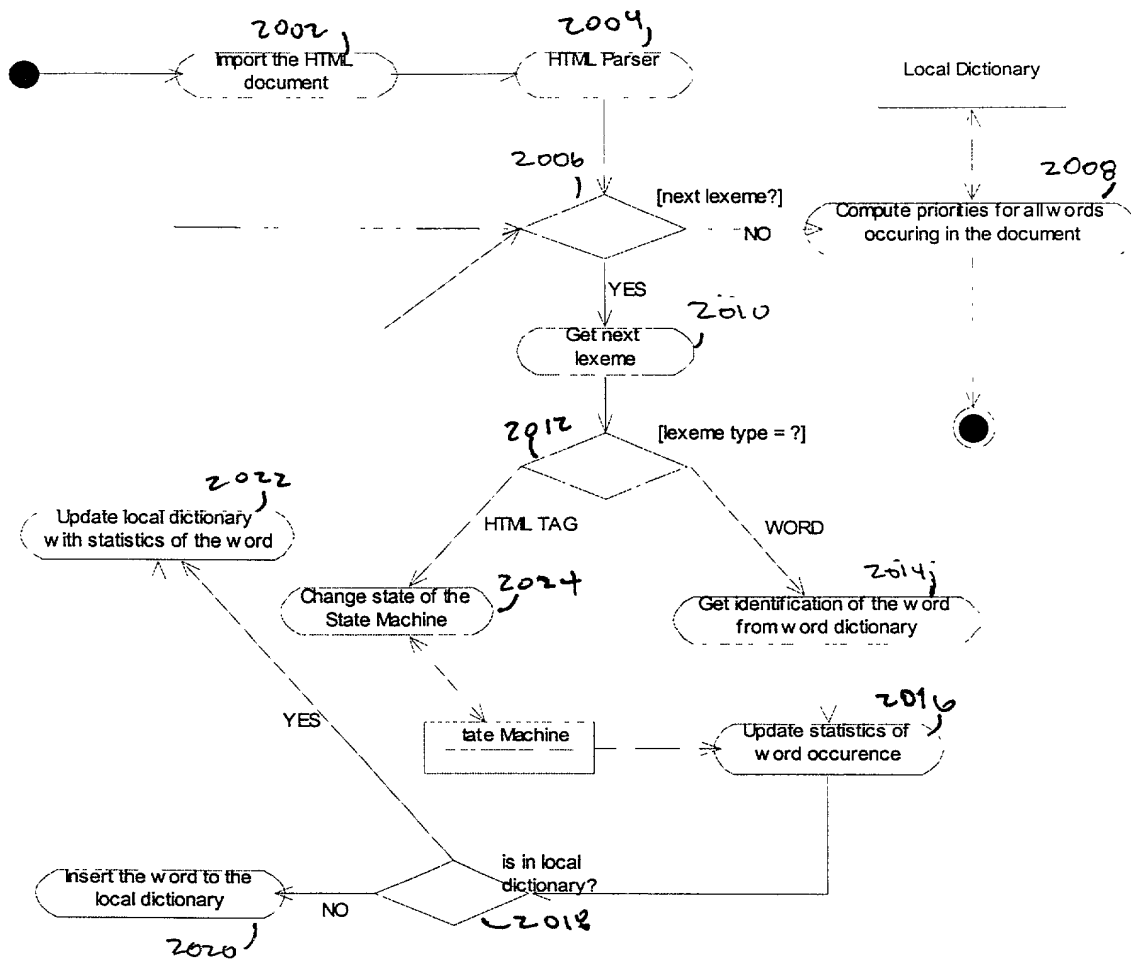


Figure 20

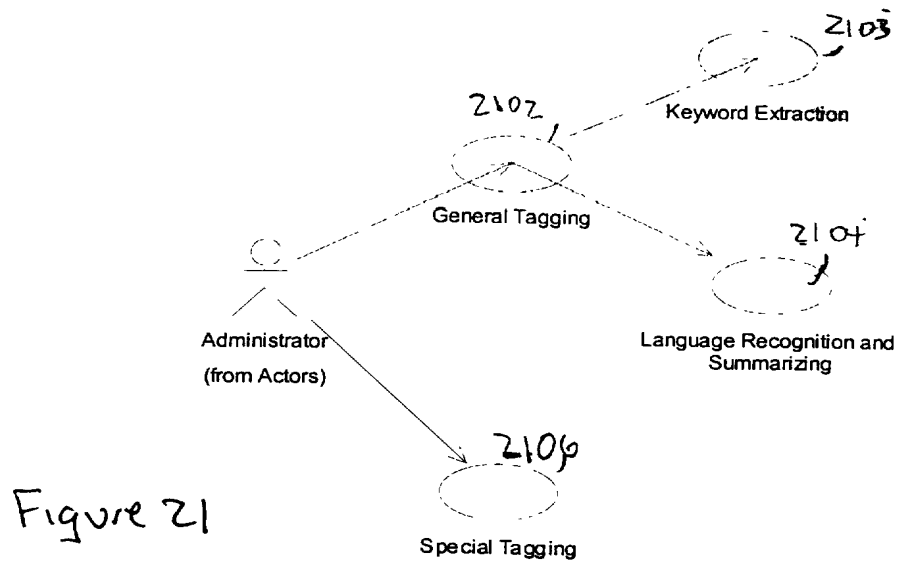


Figure 21

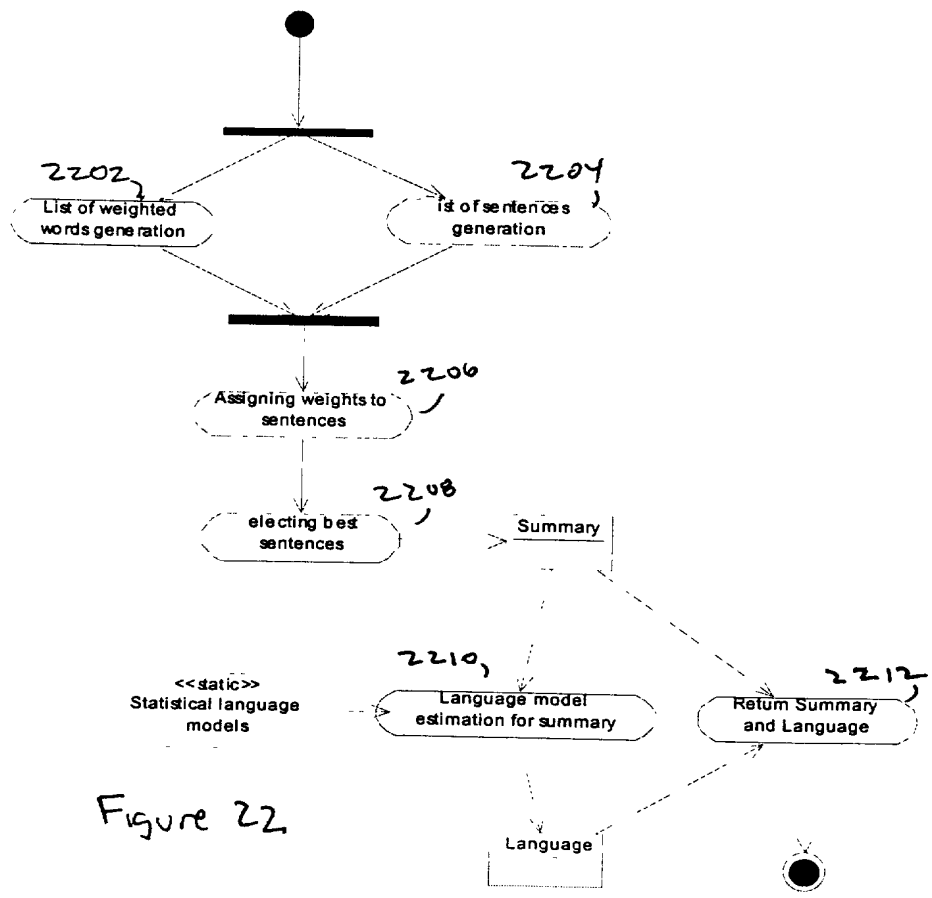


Figure 22

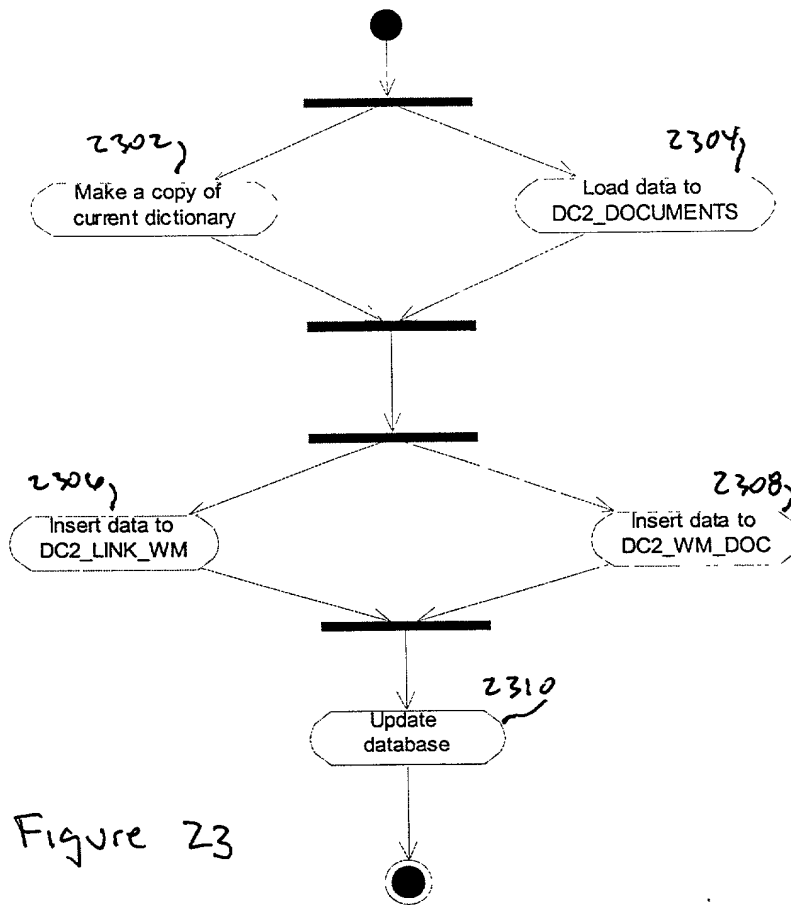


Figure 23

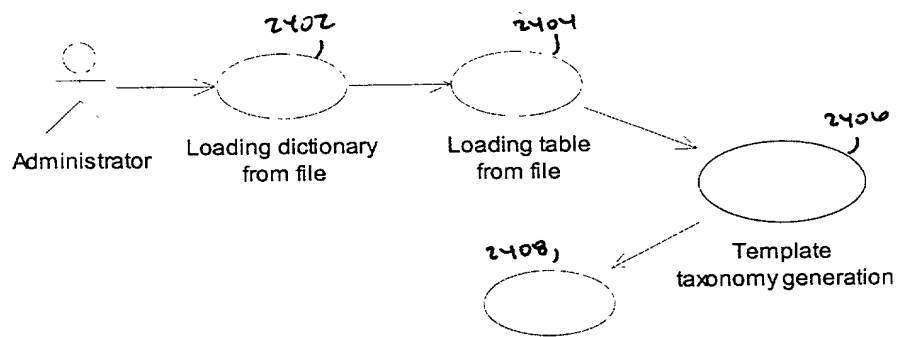


Figure 24

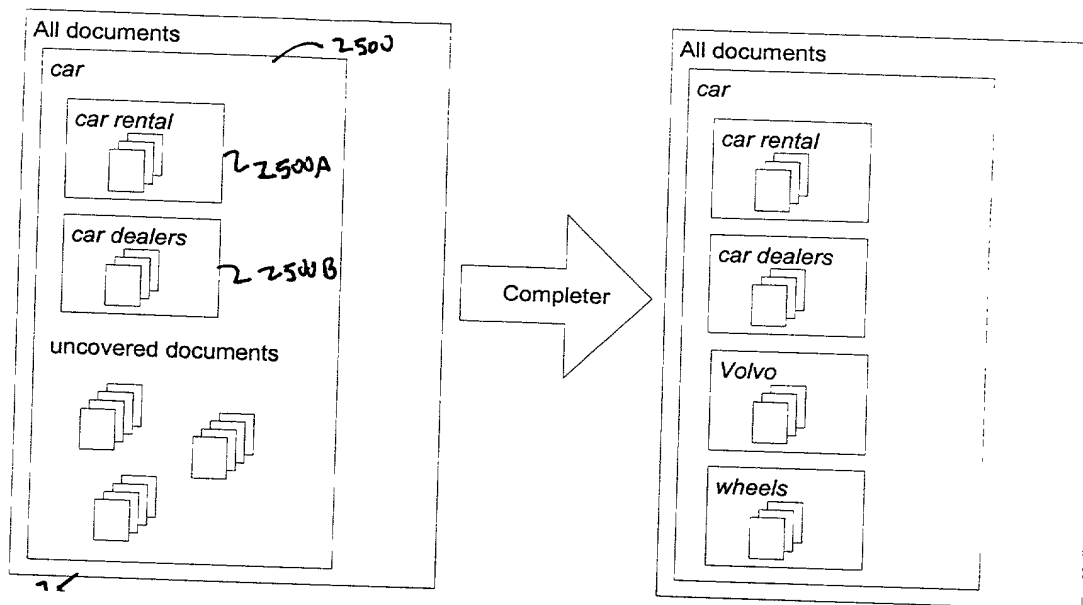


Figure 25

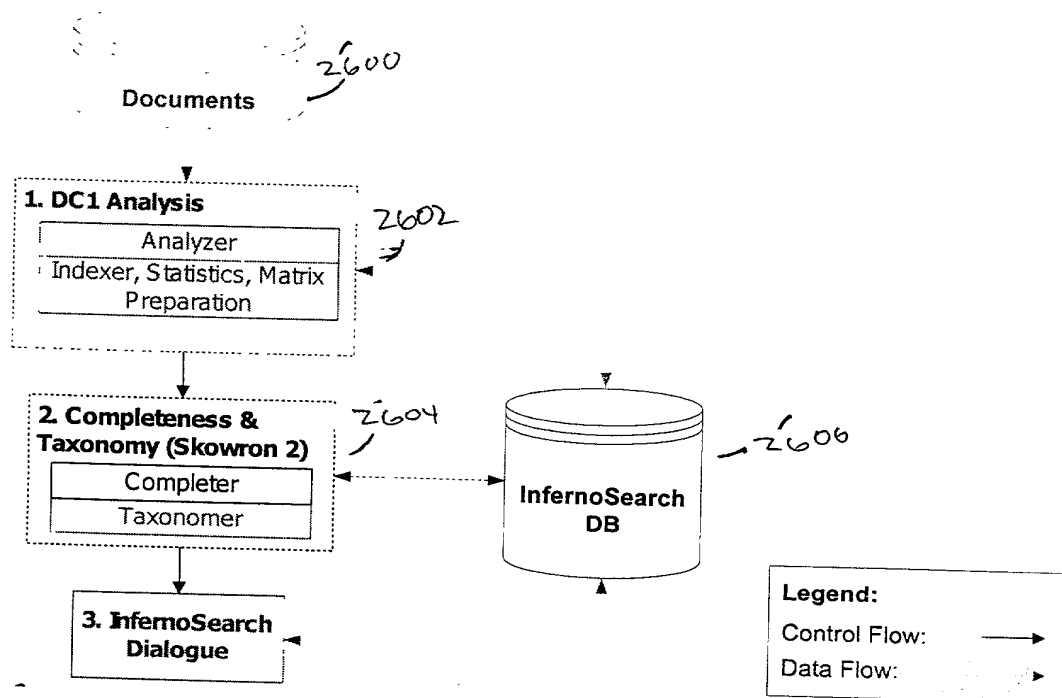


Figure 26

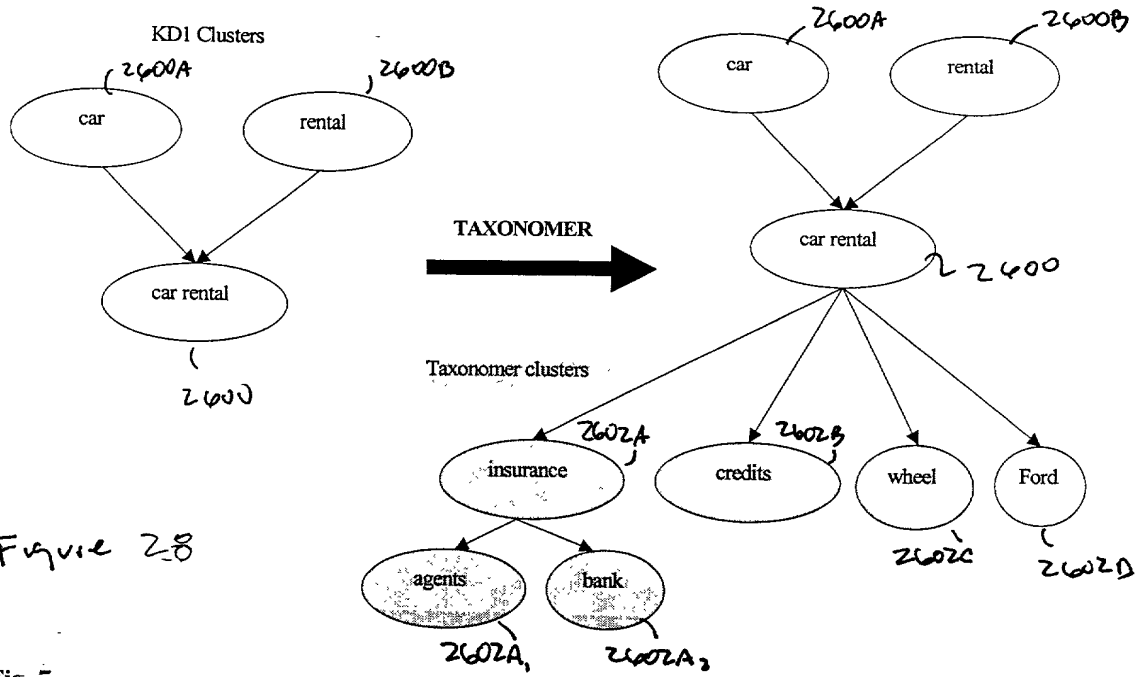


Figure 28

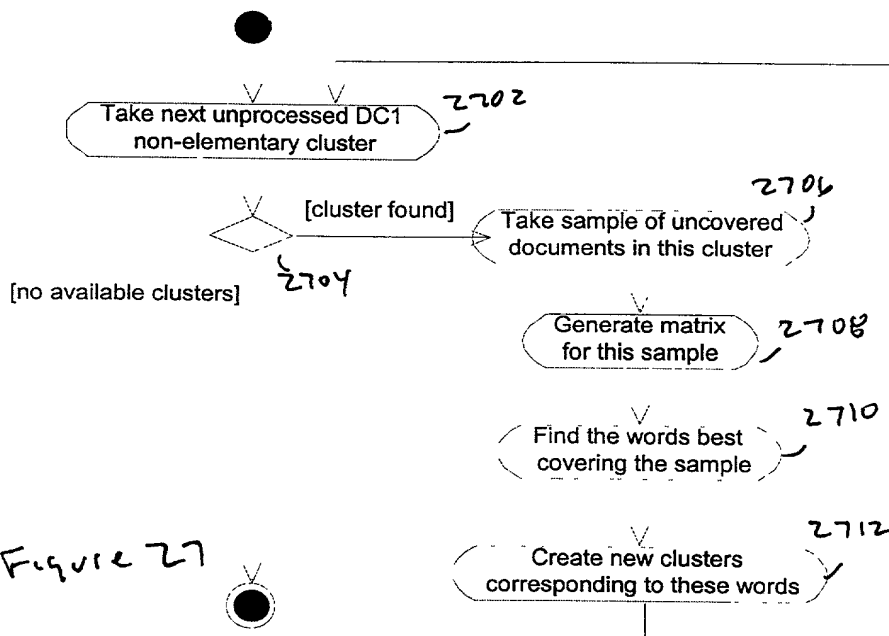


Figure 27

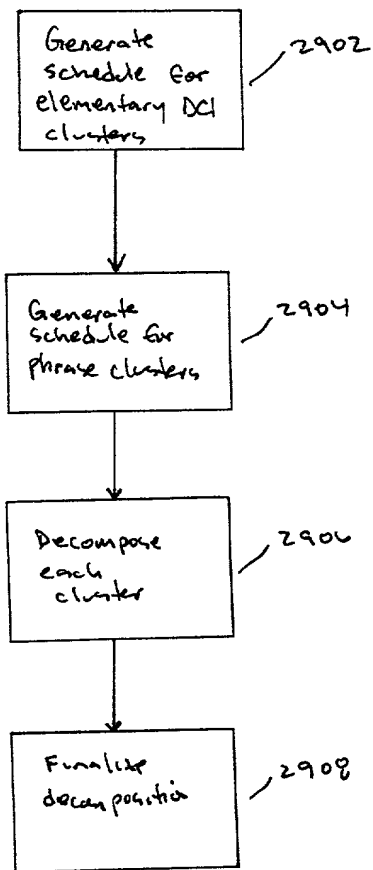


Figure 29

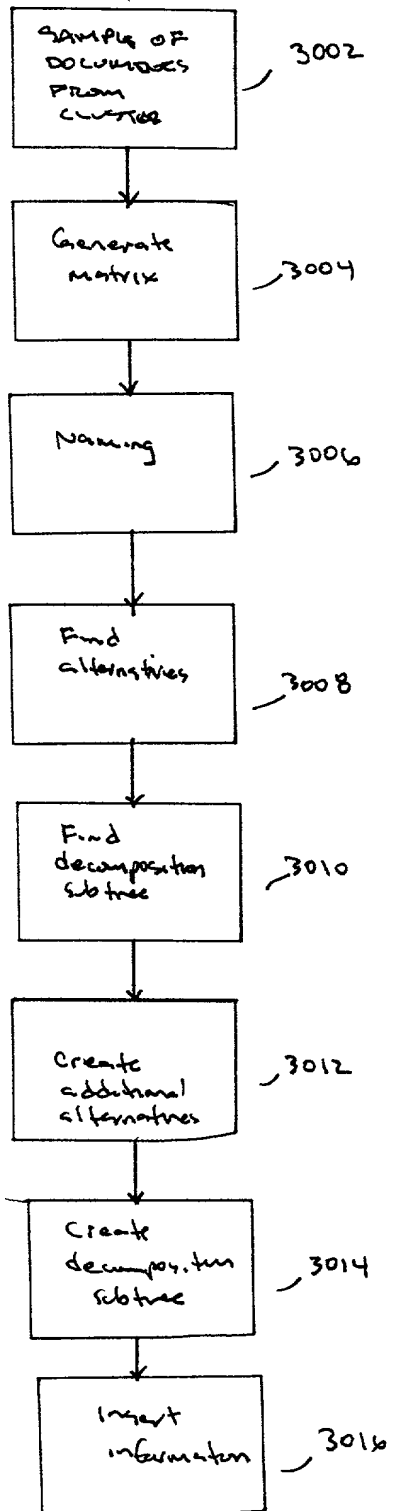


Figure 30